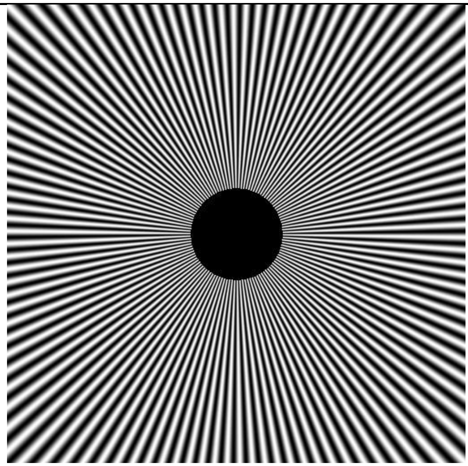
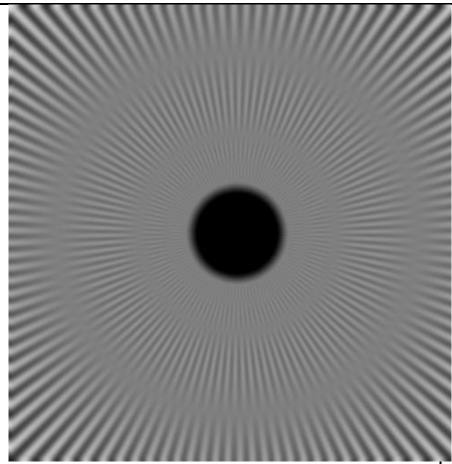
Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
ORG_A/ 1				I would like to point out an important additional test that could be included in iris camera qualification, given that your proposed test patterns include Siemens stars (or their square-wave equivalent in Figure 4, first two examples). I refer to the phase shift in the optical system's spatial frequency response. As you know, mainstream deployed iris recognition systems encode iris texture as phase bits, leading to fast and efficient matching based on Hamming distance. An iris camera with a typical aperture has an optical transfer function which, when defocused, becomes oscillatory and negative at higher spatial frequencies. This produces phase reversals if the blur circle is large enough. For typical iris cameras this effect would begin if the blur circle is about 3 pixels or larger. Phase reversals translate directly into flipped bits and elevated Hamming distances. Attached below are two images illustrating this effect: a Siemens star test pattern, and the phase shifting consequence of defocus when the blur circle diameter is about 6 pixels (demo by Udo). You can see how there are periodic phase reversals as you move radially in to the higher spatial frequencies.	The IDQT test protocol should perhaps include this use of Siemens star patterns to test for such phase reversals, as a critical criterion for how much defocus is tolerated. The actual blur circle limit will depend on several parameters of the acquisition system.	**Partial Accept** The existing test does incorporate a radial star pattern similar to the Seimen's star. The presented tests and criteria contained in the IDQT document, although not explicitly measured, are sensitive to the possible detriments from changes to the phase transfer function (PTF) as well as that of the modulation transfer function (MTF). The current incarnation of the test uses Siemens-like radial star contrast patterns (as suggested by the commenter) that can detect sign changes if they occur within the spatial frequency range for the given target. An additional sign registration marker will be added to unequivocally determine the sign of the star pattern in the object plane. In addition, the "bottom line" test, which incorporates iris-like texture patterns matching to standard-paradigm bit encoded phase features, would be sensitive to the negative impact of a phase shifts in the PTF in a similar way that an iris code would be. The suggestion to create a criterion specific for defocus is useful if the test were expanded to include aspects beyond the "peak" performance.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Sectio n/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com- ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						However, this suggestion is not in line with the IDQT philosophy of measuring the “peak” image performance of a given device, unless it is expected that a significant fraction of devices would claim defocus to be a part of their intended “peak” imaging output. **ACTIONS** • Add sign registration pattern to star target to detect phase shifting over global pattern • Reject a test explicitly measuring defocus
ORG_B /1				I know of one deployment where a camera manufacturer released a camera with a bug in the firmware resulting in all of the iris images being mirrored.	The test patterns should include tests to validate there is no mirroring or horizontal/vertical flipping of the image. For a dual eye camera, the test pattern should be able to validate that images come from the correct eye and the top left pixel is in the correct place (for instance, not the top right as in the mirrored case).	**Accept** The editor accepts the suggestion to implement such a test, as it is important for interoperability between devices and is not included explicitly in the test. A proposed solution is to indicate left and right eye markers on the face mount itself, either in the eyebrow or the eye socket region.
ORG_B /2				Ensure a continuous distribution of pixel values such that there aren’t grayscale values with disproportionally few pixel counts when the camera is presented with a gradient image. I’ve also seen this in the field where there were a few pixel values with no pixel count due to the camera’s contrast stretching.	The comment suggest that a target with a gradient of reflectivity values be included in the test to measure the albedo sampling rate / resolution.	**Reject** The four quadrant target suggested in the current version of the IDQT is used to measure both the linearity of the response of the captured images across the albedo range relevant for iris biometrics and the albedo resolution. To note, there are no explicit

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						qualification criteria based on this metric but the information can be used to develop explanations for the root cause of possible qualification failures. It may be worth considering at a later time if it is found that the linearity measurements need more than just four points within the range of iris texture to determine the linearity and albedo resolution.
ORG_C/ 1	Line 330			The exact mix of spectral composition is a trade secret that cannot be shared. It is part of each vendors R&D to capture the “best” iris image.		**Partial Accept** *(will not publicize, but will still measure)* The topic of wavelength characterization was considered at the workshop. The following points were discussed: • The primary motivation for the characterization is in the name of interoperability, following the NIST Special Publication (500-280) on *MobileID Device Best Practice Recommendations* and the draft version of the ISO/IEC 29794-6 which makes a specific recommendation regarding operational wavelength of iris devices. However, it was acknowledged that there are not studies which back these specific guidelines. The current draft version does not use the characterization in formulating the qualification criteria.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						• Considering there are no comprehensive studies reviewing the wavelength dependence of matching accuracy with the latest commercial matching algorithms within the range used by commercial iris cameras (700nm-900nm), it was suggested that before such a quantitative recommendation is used in a qualification criteria, that such a statement is backed by such a study. A dataset has been identified to which a study may be conducted. • It was noted that there are studies with some evidence showing that matching performance decreases with wavelength changes on the order of 100nm (e.g. Ngo et al 2009), with one algorithm, but this is insufficient to make a specific quantitative recommendation. • It would be expensive and impractical to perform a data collection exploring all conceivable combinations of illumination profiles between700-900 considering multiple components with varying bandwidths. **Action:** Keep characterization in as a measured quantity, but keep it out of inclusion as a part of qualification criteria. Any wavelength characterization results will not be

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of comment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						made public. A wavelength-based qualification criteria should only be initiated after an extensive study is performed exploring the wavelength dependent performance and interoperability over a number of commercial algorithms.
ORG_C/ 2	Line 385			The IPD distance of your 3D printed model should not be fixed at 63mm. You need to make different models with variable IPD between 55mm and 75mm. This accounts better for real life situations and takes into consideration gender and ethnic differences to make the test more realistic.		**Reject** It is acknowledged that the IPD can potentially be a source of image capture failure/degradation; however, the extent of the failure would be application dependent and arguably more suited to field trials or pilot studies with real people. It is the editor's contention that there are other aspects of the face model which may influence image capture and quality and inclusion at the IDQT stage of testing would require representation over relevant parameter ranges, such as skin tone and surface reflectivity, as well as other morphological features such as the eye socket and nose topology. To cover these parameters adequately would complicate the test and make it prohibitively expensive. That said, if there are particular concerns regarding known catastrophic failure on either the lower or upper 1-2% of the IPD distribution (that the IDQT developers are not aware of), and

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						would prevent a measure of "peak" performance, the information would be taken into consideration.
ORG_C/ 3	Line 846			You state that you have attempted to capture the combined influence of all the potentially significant aspects of Iris image quality yet you have not mentioned anything about assessing Focus value of acquired images? Surely you must agree that it has important bearing on the quality assessment of any imager?		**Response** The IDQT MTF and iris texture target tests are sensitive to levels of device-specific defocus which may impact the matching performance of commercial algorithms. There are no IDQT metrics that measure focus terms separate from the other important factors which may also prevent the iris information used by matching algorithms from passing through the optical system. We have chosen to measure MTF over a more specific metric which would just measure focus because other aberrations, such as astigmatism, coma, etc., could also limit the MTF of the system, not to mention other noise sources. Along the lines of this comment, there is a relevant question which is how the IDQT insures that the imaging targets are positioned or of a good nature to interact with the device so they produce optimally focused images. Although there has been effort to avoid biases, there may be certain aspects of the IDQT models which may result in a systematic offset in focus for some capture devices. A possible way around

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of comment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						this would be to include a small number of well characterized human references to confirm that systematic offsets from human targets and IDQT targets are not taking place. This issue is open for discussion.
ORG_C/ 4	Line 428			The Iris-like feature spectrum pattern you intend to generate from the BATH University iris image database will be limited by the Imager systems used to capture the BATH images. So in concept, if a new imager is better than all the imagers used in BATH database collection that new imager will not be fairly ranked as it will be limited to “seeing” only what those Imagers managed to “see” and any extra capabilities the new imager has will not be properly assessed. Another related point: you are favoring the Imagers that were used to collect the BATH database (or the newer version from same vendor) Another related point: some of the BATH images are taken using non-interlaced technology? How does that factor in a modern assessment carried out today where all imagers are digital? You need to include samples taken of REAL humans from every device that will participate to offset this bias towards those vendors whose Imager systems were used in the BATH database collection. Furthermore, what is your rationalization in selecting the area of 2.5x2.5mm for your Fourier area from which to generate the Iris Feature Spectrum? Again, you are already limited by the Imager systems used to collect BATH dataset which are older imagers and now you are potentially further weakening the Iris Feature pattern by		**Response** There may be some confusion on the presented power spectrum characterization study. The power spectrum analysis was undertaken to characterize the general relationship between a feature size, and feature contrast for human iris texture to more accurately portray real iris signals in the IDQT tests. The goal of this analysis was to estimate these general characteristics of the iris with a publicly available dataset so others may confirm the results. A similar measurement of the power spectrum using a NIR modified large format DSLR confirmed the power spectrum analysis results on the larger Bath dataset. We have confidence that the feature spectrum reproduced in the iris texture target is not significantly biased in the range of spatial frequency spanning from 1lp/mm to 4lp/mm which corresponds to frequencies near the low frequency limit of the 2.5mmx2.5mm region up to a sample rate of around 10 times that

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				making the area too large, why don't you make the area 1x1mm?		of the Nyquist sample rate respectively. Choosing a region size of 1mmx1mm would limit measurements on the low frequency end of the feature spectrum.
ORG_C/ 5	Line 707			In the iris to signal noise test you propose to compare the collected images of your Iris Feature Spectrum pattern and compare them against a "Pristine" reference template; this pauses many questions: a. Who decides what a "Pristine" reference template is?		**Response** The definition of the "pristine" template is a digital image of the synthetic iris texture target; the "pristine" templates are formed from this noiseless (by definition) digital image that is used in the creation of the iris texture target. They are independent of any device. The printed targets are however validated using images of the target taken with a large format NIR modified DSLR/lens combo with a calibrated/insignificant field distortion. These images are passed through the IDQT encoder library and compared to the pristine templates using a global Hamming distance with a nominal definition of a weak signal bit mask to arrive at the output score. There are also a templates generated just by the device for the "instrument only" comparisons. Special considerations are given in the case of significant discrepancy between the scores resulting from instrument only and pristine comparisons.
ORG_C/ 5	Line 707			b. Which algorithm will you be using to generate the template of this "Pristine" image? And why did		**Response** There are no commercial algorithms

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1 Organization /Comment ID	2 Section/ line #	(3) Paragraph/ Figure/Table/ Note (e.g. Table 1)	4 Type of com-ment[2]	5 Comment (justification for change) by the organization	(6) Proposed change by the organization	(7) Proposed Editors Disposition
				you choose that algorithm not some other algorithm? Will you consider multiple “Pristine” templates from multiple algorithms just to avoid the known biases that certain algorithms have for certain types of imagers and to make sure that imagers work well across not only different ambient environments but work well with different algorithms?		used in this test. The algorithms that are used follow the general paradigm of a binary feature encoder matching algorithm, using 2D gradient and ridge filters at different orientations in the pseudo-polar coordinate system. A number of filter basis sets have been explored, including log-Gabor, Haar, DCT, and FFT forms. Pending the comments and feedback from the workshop, the IDQT plans to use one form, as the initial development exploration revealed similar results for each form. The purpose of using an incarnation of the encoder/matcher paradigm is to attempt to get closer to a definition of iris signal in which to develop a signal-to-noise statement, but without using a specific definition from a proprietary commercial algorithm. Also, the results from commercial algorithms cannot reveal performance as a function of the individual spatial frequency bands which is required to form the IDQT metrics. A discussion on possible biases or the effectiveness of such a method would be welcome, considering this is considered a “bottom line” test which has an important role in assigning qualification criteria.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
ORG_C/ 5	Line 707			c. At this stage, our comments on line 428 takes on a more pressing note, if the core image you are asking every imager to capture (Iris Feature Spectrum) is generated from examining a 2.5x2.5cm area from each of the BATH University database of images and those images are biased towards a certain vendor or a certain accuracy level (weak images with low quality/clarity/focus/illumination and motion blur) and based on that you are generating a pattern against which newer imagers will be evaluated? This has the side-effect that you will not be able to distinguish imagers that are much better than the ones used to collect the BATH sample from which you are generating your "perfect" Iris Feature Spectrum pattern. You need to offset this by including images that are collected using different imagers in the set from which you intend to generate your Iris Feature Spectrum image.		**Response** The intent of the iris texture characterization was to measure the intrinsic albedo variations of the iris pattern as a function of spatial frequency. There was an assumption that the Bath dataset had a pixel values that had a linear response and so, although they did not have a direct albedo calibration, they could be used to estimate a functional form of the feature contrast with spatial frequency bootstrapping the absolute scale to average albedo values. The iris texture characterization on the Bath dataset was checked with a smaller number of NIR images taken with a large format NIR modified DSLR that was calibrated to albedo values. Also, the spatial frequencies measured were well sampled on the Bath dataset in a range where the spatial frequency response was near unity. There is confidence that the albedo feature characterization, over the range of interest, should be independent of device. The iris feature spectrum based on this calibration then should also be largely independent of a given imaging device.
ORG_C/ 5	Line 707			d. We are concerned if there will be only one Iris Feature Spectrum pattern and template,		**Accept** There may be merit to including more

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				perhaps you should deviate your generation algorithm to generate 20 base images from the BATH database of images just to avoid the possibility of having a single image that is not correct, if that happens you have no way of detecting it.		than one iris texture pattern in the test in the name of stochastic averaging. 20 patterns however may be in the arena of diminishing returns and would amount to a much larger data collection effort for each test. Although we want to keep the test as simple and efficient as possible, it may be worth considering 3-5 different patterns, so a variance can be established. We would then either justify moving back to just 1 sample, or expanding to larger numbers. This should be discussed at the meeting.
ORG_C/ 6	Line 739			You propose to measure the Greyscale Linearity and illumination Uniformity as if the targets are real human beings with real irises. The ink-printed targets will behave differently for sure. We caution against over interpretation of the findings here as human tissue will behave and react differently under NIR than Ink.		**Noted** We have undertaken a thorough characterization of a number of commercially available inkjet inks and have found a set which has suitable albedo characteristics in NIR wavelengths to reproduce iris-like features in the spatial frequency ranges of interest in IDQT. We do not claim to have a set which reproduces the albedo characteristics of the human iris for wavelengths outside of the range used in iris biometrics, but this is arguably not necessary for the IDQT. The possible contrast changes with illumination angles having to do with the 3D structure of the iris would not be included in the current test. If

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						aspects required for image capture of a particular commercial device which are not included in the test, the IDQT development effort would appreciate feedback in the spirit of minimizing bias in the testing process. For example, if iris image information as seen in visible wavelengths is used in the capture process for a particular device, a note from the device manufacturer to the test would help minimize bias.
ORG_C/ 7	Line 776			Once again, you mention two templates that you intend to generate, a “perfect” reference template and an Instrument Only template. What do you mean Instrument only template? Also which imager will you use to capture the “perfect” reference image with? And will that same imager ever undergo IDQT certification? That would be very unfair to other imagers? Finally, what algorithm will you use to generate the “perfect” template and the “instrument only” template and we are assuming that you will be using the same algorithm to do the matching as well.		**Response** There is a possibility that for some devices, images taken with the evaluation device of the iris texture target may match well to each other, but not as well back to the “pristine” template defined in albedo space. There are a number of reasons why this may be the case. The instrument-only template comparisons are used to determine to what extent the evaluation device is of this nature. If the match scores from instrument-only comparisons are significantly lower than that compared to the “pristine” template, then this is an indication that the device may be capable of recording iris information, but may have issues concerning interoperability. If a significant number of devices exhibit a lack of compatibility with the pristine

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Sectio n/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						template, but perform well in the instrument-only comparisons, then the test may need to consider modifying the "pristine" template.
ORG_C/ 8	Line 799			You state that your qualification criteria is designed to be algorithm agnostic? How do you propose to achieve that when you are using "an" algorithm to perform the matching and to produce the HD upon which the classification of the three levels is built? While applying a single algorithm across board gives you uniformity, your approach will be influenced by the particularities of that single algorithm and its own internal preference to illumination, brightness, focus etc. We strongly propose you repeat the same Level testing using at least three different algorithms to avoid this clear weakness in the test.	Validate the three level tests	**Partial Accept** The goal of the test was to be algorithm agnostic, and by "algorithm" it is meant commercial iris biometric matching algorithm. Attempts have been made within IDQT to characterize and to replicate aspects of the human eye that are intrinsic to the signals used in commercial iris biometrics. If this is done adequately, the test should largely not favor one algorithm over the other. It is certainly possible that the definition and characterization of the human iris used to formulate the definition of the IDQT iris targets used in the test do not fully encompass the definition used in commercial matching algorithms. For example, if some algorithms incorporate features found in the periocular region, or some other aspect of possible information contained in an iris image but not included in the IDQT, then there would be a case for algorithm bias. Text will be changed to state that attempts have been made to make the IDQT matching algorithm agnostic, but it is admitted that this has not been

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						confirmed, or to what degree it effectively evaluates device image quality in a way which equally encompasses all commercially available matching algorithms. Test target validation image targets are compatible with a few leading commercial iris matching algorithms
ORG_C/ 9	Line 853			You have three lighting conditions, indoors, indoors with windows and outdoors. We are not clear what you will be testing in the outdoors; for example there are imagers on the market that will work outdoors because they are designed with a hood to obscure the outside ambient lights. Will this mean that imagers that do not have such a plastic (and relatively insignificant solution,) will fail your test? Are you testing how a vendor could “cover/shield” their imager using a plastic hoodie or are you testing the effectiveness of the built-in NIR filters in blocking the sunlight? Please clarify what you will be testing in the outdoor exactly taking into consideration that whatever filter is used and at whichever intensity level it will be overwhelmed by the sun’s powerful transmissions in the same wavelengths since the sun (as you know) has all wavelengths. Basically there is no way to block the sun without an add-on cover or shield that physically shades the target. We have deployed our AD100 unit and other units on ATMs in the streets, but we do require that the ATM has at least a “cover” and not be in the direct sunlight. If we are to make it work in direct sunlight we could design a plastic hoodie and ask people to stick their heads inside it and we will be done, please clarify what will you be testing in the outdoor		**Response** The outdoor ambient lighting test uses a patterned scene with an integrated illumanance level which is typical of that which is measured outdoors (solar-like spectrum) in sunny, or partly cloudy conditions. The purpose of this test is to qualify devices as they stand by themselves without other external mitigation techniques to shade the scene from the eye to determine if the device as submitted would be suitable for outdoor use. It is acknowledged that placing a piece of plastic or other opaque material to block the scene will likely mitigate the influences of ambient lighting on image quality, but this also may introduce a contact aspect of the device. An explicit contact/non-contact category should be included in the device type.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				situation.		
ORG_C/ 10	Line 884			When will the tests be taking place? Will this happen in one batch for all vendors? Will you repeat the testing every year? Will compliance with this test be mandatory for selling to Iris imagers to DHS and other US-government agencies? Will the results of all testing of all imagers be published for everyone to see or will the results be private and only provided to vendors?		**Response** These are valid questions but beyond the scope of this document. For the near future the tests will be conducted at the discretion of the supporting US government entity, namely the DHS. IDQT may be used in the future for additional applications or by other organizations.
ORG_D/ 1				A bit surprised, and quite pleased, to see the concept of 1:1 in this document. I think we both realize bigger challenge awaits downstream with users selecting the right device for the application at hand. It's encouraging to see the concept has been addressed in the draft.		**Noted** The editor is grateful for this supportive comment, and agrees that the application specific tests involving humans in the loop presents perhaps bigger challenges.
ORG_E/ 1	138-141			It is a good goal to remove human frailty from iris camera testing by measuring 'peak' imaging capability. However, if operator-use produces substantially poorer images as a result of device design, 'peak' results might be misleading. For example, if a fixed focus camera operates with a shallow depth of field (e.g., with large aperture) and, as a result, produces admirable spatial resolution when focused perfectly, it might also produce a distribution of images when operated in the field that on average shows substantially poorer spatial resolution. In contrast, an autofocus camera designed with a greater depth of field might have a poorer 'peak' spatial resolution but might produce a distribution of images with spatial resolution when used in the field that is better than the fixed focus camera that relies on human operator and subject behavior. Making this	Add further explanation in the document regarding the implications of the peak imaging philosophy, and add clarification text in the document**.**	**Response** At the most basic level, the IDQT is designed to answer the question, "removed from human-subject interactions, is the device under evaluation capable of recording the information used in iris biometrics?", with issues of human interaction left for the next stages of testing. This motivates the IDQT philosophy of measuring the "peak" imaging performance of iris cameras removed from such human interaction aspects as motion, control of pupil dilation, eye gaze, and occlusion. After considering all possible aspects of a device that

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				distinction clear to users of the IDQT is essential to the goal of IDQT. As pointed out in the July 10 meeting, inclusion of the Capture Volume (defined as the volume in space inside of which a properly captured iris can be matched to a given match score threshold) will complement the peak image information and inform the reader of the image quality/depth of field trade-off. Further matters concerning peak and typical image quality can be reserved for human-in-the-loop testing.		could influence image quality related to human behaviour or interaction devising laboratory tests without involving live human subjects would either be overly costly, or inaccurate. For example, devising a laboratory test to measure how a device effectively controls eye gaze, without a human in the loop would arguably be ineffective without having a human subjects involved. For fixed focus devices, this would be revealed in the claim of capture volume and verified in the IDQT. For systems designed to work with moving human subjects walking through a relatively tight depth of field, the peak imaging performance would be measured on at the optimal focus of the device.
ORG_E/ 2	Lines 146-149			Anticipating the need for greater information content is forward-looking and cannot be faulted. However, the notion that there is a 'best' performance band (level III) will give prospective clients the impression that cameras in lower bands (e.g., level II) are inferior. The result might delay or dissuade a client from purchasing a (level II) device in anticipation of better (level III) performance. Such client behavior has a precedent in early versions of ISO/IEC 19794, Annex A which differentiated spatial resolution (pixels/iris diameter) with 'marginal', 'acceptable' and 'good' bands without supporting data. ORG_E recommends clear explanation with the text of IDQT that more iris information will have impact on cost,		**Reject** **Clarification:** The intention in the tri-level testing is not to define a "good, better, best" type scale, but rather to distinguish the ability of a given device to pass different spatial frequency information that may be useful in identification of an individual. As stated in the text, "the exact qualification criteria used in a given project would likely depend on the application." For example, if a level I criteria is deemed all that is needed for a given

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				complexity and performance of iris cameras but might not correlate to better biometric performance under all conops. Perhaps renaming the performance bands as Low-Cost (Level I) and Standard (Level II) might be more suggestive. A name for Level III needs to suggest high resolution accompanied by higher cost and complexity or higher constraint (e.g., closer subject distance.)		application requirements, then it would be included for further human-in-the-loop evaluations along with those that may have passed at a level II or III. If the performance in the human-in-the-loop tests of the level I device was matched that of level II and level III devices, and was less expensive than it would likely be viewed as more favorable compared to the level II and level III devices. Therefore, there should be no misconceptions that level I results are inferior to level III from the point of view of the final assessment in the larger scale evaluation process. There is a question, out of scope for the IDQT, but still important which is how will level I, II, and III application requirements be determined… The suggestion to associate level I with low-cost would be presumptive as it is certainly possible for a high cost, level I device to exist. The following sentence was added after line 149: … “and to be clear the different levels of qualification criteria do not represent absolute metrics for procurement decisions which would take into account many other factors such as cost and the performance measured in human-in-the-loop evaluations. They should be viewed

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						rather as an important framework to effectively formulate requirements that are appropriately matched to a specific application.”
ORG_E/ 3	Lines 160-161			ORG_E recommends noting that testing on moving targets is out of scope of the current IDQT and therefore that IDQT cannot aid in evaluating cameras that claim to capture iris imagery on moving subjects. (If IDQT testing will include targets on rails that can accommodate motion as suggested at 7/9 meeting, this should be added as a footnote.)		**Response** It is true that the test has not been developed to compare performance as a function of subject motion. However, in a further section on lines 605-606 it is mentioned that “improvised solutions will be employed with feedback from the vendor” for devices that require subject motion for capture. Ideas to be tried involve mounting the target assembly on a rail with an open degree of freedom along the optical axis and initiating motion as prescribed by vendor instructions. It is not explicitly stated that the IDQT does not test performance as a function of subject motion, and this will be clarified by the following addition to the text at the end of the paragraph on line 164: “Although the IDQT is designed to accommodate different capture modes, it should be noted that the test is not designed to explicitly evaluate devices on their ability to capture moving subjects.” A possible solution to mount the target on optical rails is under consideration.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
ORG_E/ 4	Line 169			Use of the word 'accuracy' is open to wide interpretation. On one hand, images from high speed iris cameras are not as pretty as many single image cameras. On the other hand, biometric statistics using high speed image acquisition might not suffer in terms of accuracy since multiple comparisons to a reference image are made in the course of a single subject-transaction. Therefore, ORG_E recommends replacing of the word 'accuracy' with 'single image quality.'		**Accept** The word "accuracy" is replaced with "single image quality" in line 169.
ORG_E/ 5	174-175			ORG_E commends IDQT on efforts to remove bias due to single vs. multiple image capture.		The editor appreciates the positive comment**.**
ORG_E/ 6	182			10 seconds might be acceptable in a test environment but ORG_E recommends 3 seconds for a time-out in practice as well as making the distinction between testing and practice clearer in the text.	Add clarification text	**Accept** The timeout period was chosen not to reflect an application requirement which would likely be much shorter (more like the 3 seconds suggested), but to provide more than adequate time to give the opportunity to devices to deliver their "peak" imaging performance. The following clarification text is added after line 182. "It should be noted that the timeout period used in the IDQT does not reflect any application requirements, or would be considered for use in any "best practices" recommendation for the operational use of iris devices."
ORG_E/ 9	241-242			Indeed, it might be argued that there need be no difference in quality between enrollment and verification (probe) images for certain iris recognition applications since matching uses information that the reference (enrollment) and probe images have in		**Response** In early deliberations of the IDQT, there were discussions on defining how one may define an enrollment image. This section was created to clear up the

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of comment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				common. However, this argument is unresolved at present so it is good that IDQT does not fix a definition for enrollment vs. verification image quality. Therefore, why include section 2.5?		misconception. In particular, it is the opinion of the editor that a universal definition of an enrollment image could be defined and that the IDQT was a good place to define it. It was kept in place to discuss the position on the issue of defining an enrollment versus verification image, in that it should be application dependent.
ORG_E/ 10	258			Need to include Appendix explaining spatial sampling rate, Nyquist sampling rate and assumptions regarding point spread function such as connection of sampling rate and MTF. Current manuscript lists 'appendix xxx' which is not included.		**Accept** The revised edition of the document will include the appendix referenced in the text.
ORG_E/ 13	299			See B. Clark et al., Am J Optom Arch Am Acad Optom. 1971 Apr; 48(4): 333-343, which suggests corneal reflectivity of 8%.		**Noted** The editor is grateful for this reference. Although the 2-3% level was stated in the text, fortunately in the target development, the artificial cornea reflectivity was calibrated using the reflection from real human eyes, not to an absolute level. We are confident that the test will depict how a human cornea may reflect incident ambient light to a level necessary to formulate a comparative evaluation. This reference has been stated in the text.
ORG_E/ 14	324-332			ORG_E agrees with the editor's comments – no evidence connects the draft ISO/IEC 29794-6 illumination spectrum requirement to biometric performance. While collecting such information is useful, especially in the context of the multi-spectral work in progress and		**Partial Accept** *(will not publicize, but will still measure)* The topic of wavelength characterization was considered at the workshop. The following points were

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				work done previously, ORG_E questions the use of evaluating illumination spectrum at this point and recommends suspending this test until a connection to biometric iris recognition statistics is demonstrated. ORG_E further recommends not reporting the results of any spectral testing at this stage, awaiting results of ongoing studies.		discussed: • The primary motivation for the characterization is in the name of interoperability, following the NIST Special Publication (500-280) on *MobileID Device Best Practice Recommendations* and the draft version of the ISO/IEC 29794-6 which makes a specific recommendation regarding operational wavelength of iris devices. However, it was acknowledged that there are not studies which back these specific guidelines. The current draft version does not use the characterization in formulating the qualification criteria. • It was noted that there are studies with some evidence showing that matching performance decreases with wavelength changes on the order of 100nm (e.g. Ngo et al 2009), with one algorithm but this is insufficient to make a specific quantitative recommendation. • Considering there are no comprehensive studies reviewing the wavelength dependence of matching accuracy with the latest commercial matching algorithms within the range used by commercial iris cameras (700nm-900nm), it was suggested that before such a quantitative

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						recommendation is used in a qualification criteria, that such a statement is backed by such a study. A dataset has been identified which a preliminary study may be conducted. • It would arguably be expensive and impractical to perform a data collection exploring all conceivable combinations of illumination profiles between 700-900, while considering multiple components with varying bandwidths. Action: Keep characterization in as a measured quantity, but keep it out of qualification criteria. Any wavelength characterization would not be made public. A study of wavelength interoperability, which include a number of commercial algorithms, may be conducted.
ORG_E/ 15	441(Fig. 4)			ORG_E recommends using 30 and 60 periods as descriptors for star pattern test targets.		**Accept** **Clarification:** The source of misunderstanding was identified. The text described the star targets in terms of the number of segments in the 360 degrees rather than the number of periods. The targets intended for use are indeed 30 and 60 period targets. The radial targets are chosen to provide high signal-to-noise measurements of MTF at spatial frequencies at 1, 2 and 3 lp/mm, and to modulate the radial location of where specific frequencies

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of comment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						are made. To note, the outer radial regions of the target contain the most area to make MTF measurements with star targets and result in measurements with more confidence.
ORG_E/ 16	492-495			Handheld iris recognition devices that are intended to be small but also are designed for distant iris capture beyond 1m typically produce retinal retroreflections (infrared-eye) because of the small angle between the camera-pupil and illuminator-pupil axes. Tolerance of non-black pupils is built into some algorithms today and will be designed into more algorithms in the future. Therefore ORG_E recommends that future test design not only tolerate but characterize retinal retroreflection.		**Accept** This is a relevant point. It was decided that this test not be included in this version of the IDQT due to the complexity and cost in constructing a realistic human eye target that also had reproduces realistic retinal retroreflection. The IDQT effort is accepting any suggestions to practically construct such a device.
ORG_E/ 17	515			ORG_E recommends that IEC 62471 be listed explicitly as the standard for eye-safety because it is designed around LEDs rather than LEDs and lasers. The reference to lasers in the ACGIH handbook can be confusing. If laser light sources are used, ORG_E recommends characterization of accompanying speckle pattern noise.		**Accept** The eye safety guidelines used will be those most appropriate for the illumination source. If a laser source is used, a characterization of the speckle pattern will be carried out to ensure eye safety statements are taken into account; to include, the constructive interference that may produce higher irradiance values at an eye as compared to averaged values.
ORG_E/ 18	531 and 651-652			ORG_E recommends that the solar (broadband) spectrum used for ambient light qualification be listed explicitly in the form of a table and graph to allow reproducible testing at other sites.		**Accept** The editor agrees with the commenter that further clarification on the illumination source is needed in the text. The IDQT ambient lighting test will use an illumination source that is

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com- ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						similar to that of a black body spectrum, with a temperature close to that of the sun (~5800K) in the range between 700 and 900 nm only. Mitigation solutions which may use higher resolution solar spectral features (such as the terrestrial oxygen feature at ~760nm) would be negatively biased and would be handled with communication with the vendor. More specific information on the exact light source to be used for the ambient light scenario test will be included in the next draft of the IDQT document. The illumination levels that define the three ambient light levels were chosen based on a rounding of measurements with a calibrated irradiance meter to the nearest order of magnitude. This was undertaken considering the variability of scene illumination levels due to clouds and the integrated reflectivity of the scene presented to a subject undergoing iris image capture.
ORG_E/ 19	757			It is possible that corneal reflectivity is higher than predicted by an estimate based on Fresnel equations and corneal index of refraction. Again, refer to B. Clark et al., Am J Optom Arch Am Acad Optom. 1971 Apr; 48(4): 333-343.		**Accept** This is a good reference that was overlooked. The corneal reflection representations in the IDQT targets have been calibrated to a small sample of human eyes, not on an absolute level. Thus, we feel the targets are adequately representative of how the

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment²**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						cornea reflects incident light to the purpose of the IDQT.
ORG_E/ 20	778-784			Assessment of the capture volume requires biometric iris matching throughout the volume. ORG_E systematic measurement of images along lines defining the principle axes of the volume and along lines that trace the extremities of the claimed capture volumes. Biometric match scores along such lines will determine the actual capture volume. This, of course, requires a choice in match score threshold which necessarily links the algorithm to the test results. Therefore, capture volume is by nature, a compound hardware/software parameter. Furthermore, capture volume is affected by the same optical parameters that influence spatial resolution. The coupling of capture volume (depth of field) and spatial resolution at peak focus and with degrees of defocus needs to be included somehow.		**Partial Accept** A similar method of validating the manufacturer claims of capturing volume is suggested in the existing IDQT document, with the iris texture target and the three level iris feature encoding and matching algorithms. There are no intentions of using commercial matching algorithms as part of the test at this point. Instead of testing throughout the capture volume on a grid as suggested, the nominal test suggested in the IDQT would just test the outer boundaries claimed by the manufacturer.
ORG_E/ 21	822,829,837			Matching with a HD of 0.1 or less is arbitrary and suggests that a match with HD = 0.14 is somehow less informative than a match with HD = 0.09. Using strict Daugman statistical definitions, HD is a direct measure of matched information but image to image variation suggests that relative certainty based on HD is important. For example, three consecutively taken iris images might yield HD = 0.09, 0.15, 0.12, all matches but ranging by 0.06 in Hamming distance. Does this mean that the image corresponding to the 0.15 score is worse than that corresponding to the 0.09? Or are the vagaries of segmentation responsible for the variation? Because of the role of algorithm, ORG_E recommends that bare iris feature spectrum target match score threshold of 0.1 be reconsidered and suggests 0.2 or		**Partial Accept** Applied to an operational scenario where matching occurs between two of the same irises taken at different times, variables such as pupil dilation, eye gaze, and segmentation errors due to variable occlusion are examples of influences that can increase a Hamming distance score for genuine comparisons. Static tests are used in the IDQT which arguably should have criteria to represent the lowest scores possible in a realistic match distribution and should be arguably lower than matching thresholds for a real

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				0.24 instead. If there are data to support a threshold of 0.1, it should be included as an Appendix since there is no support for such a score threshold in real iris matching for the reasons listed. This goes for levels I, II and III.		application. That said, there is the very valid question of how 0.1 was chosen as the criteria over other possible values. This was chosen from a Heuristic test using a wide variety of different quality images of the target pattern with a range of illumination (photon noise) and focus (MTF) values. It was observed that a transition in the Hamming distance metrics occurred in the range between 0.07 and 0.13. A less arbitrary definition of the threshold value will be developed in the final IDQT version.
ORG_F/ 1	154-155			Other metadata could include image dimensions, format, storage/file size, image naming, EXIF, timestamp, etc. as well as things that are controlled/known like shutter speed, aperature, etc.		**Response** The input/output format for evaluated devices is noted, but there is no qualification criteria associated with this data. The suggested metadata may be more appropriate for the conformance test to ISO 19794-6 data exchange format for iris biometrics. Other data of the EXIF variety is not used in any step in the IDQT process.
ORG_F/ 2				If possible, both eye images should be collected at once and stored correctly as L, R. We had a lot of issues with certain devices at Ft Bragg, and during download of EFT files vis-à-vis ground truth		**Accept** See comment from MorphoT/1
ORG_F/ 3				Battery life may be important in some scenarios… suggest asking vendors what it is.		**Noted** Battery Life is currently not tested as a part of the IDQT, as the focus is on image quality. There is an argument

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						that this should be verified before a field test, however at this time the IDQT will not test battery life.
ORG_F/ 4				Who can see the evaluation reports besides the vendor, testers, NIST?		**Response** The official policy for handling test results is under development. The sensitivity of the test results is acknowledged and a policy to protect the information will be developed.
ORG_G/ 1				Line item heading 333 - Eye Safety, references ICNIRP Statement on LED and Laser Diodes: Implications for Hazard Assessment. This inclusion may suggest that the NIST draft document recognizes and limits diode-based illumination technology as the only acceptably safe illumination technology for iris imaging. In contrast, US patent 8,254,768 discloses an alternative iris illuminator technology that is not based upon LED technology and yet is eye safe to applicable human eye safety standards. This alternative illuminator technology solves fundamental deficiencies with LED iris illuminators and delivers greatly improved performance over traditionally challenging conditions like full sunlight. Importantly, this alternative illuminator technology is eye safe with >10X margin of safety to relevant illumination eye safety standards though the safety calculations are not specific to LED type illuminators, which the NIST draft references. The NIST Special Publication as drafted on May 8, 2013 may imply to eliminate other types of iris illuminator technologies by suggesting the illuminator must <u>only</u> be LED based technology in order to meet referenced LED eye safety document. This treatment in the NIST draft is		**Accept** The eye safety guidelines used will be those most appropriate for the illumination source. If a laser source is used, a characterization of the speckle pattern will be carried out to ensure eye safety statements will take into account the constructive interference which may produce higher irradiance values at an eye compared to averaged values.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				unnecessarily narrow in its scope and perhaps unintentionally eliminates a potentially superior technology by favoring only LED type illuminators. The conclusion to this comment that it is suggested to eliminate the narrow treatment that implies LED technology is the only eye-safe illuminator technology. Alternatively for final NIST Special Publication, the NIST document should delete ref. 4 and alternatively apply Reference 8 (ACGIH) and/or another recognized eye safety standards like IEC 62471, Photobiological Safety of Lamps and Lamp Systems. By doing so it would be inclusive of all iris illuminator technology candidates that are compliant to eye safety standards.?		
ORG_G/ 2	325-332			Illumination wavelength content. While I compliment what I believe to be the intent of this section by promoting broadband NIR illumination per ISO/IEC 29794-6 draft, the actual expression of the wavelength details is too narrowly treated for actualizing the fullest potential for maximum iris performance. In addition to NIR broadband illumination, the attached paper 'Multispectral Iris Analysis Preliminary Study', C. Boyce, et al validates that an iris imaging system benefits from other wavelength bands and yields improved iris matching performance. Because NIST Special Publication draft line items 327, 328 and 329 (echoing ISO/IEC 29794-6 draft) uses a percentage of all illumination as a bounding metric, it becomes overly constraining and thus eliminates other meaningful contribution from other beneficial bands. For example, line item 327 of the draft would limit to 10% or less of all supplied sight to be in the blue band (400~500 nm). And	The recommended for changes to lines items 327, 328 and 329 follows: *The NIR broadband content between 700 and 900 nm is recommended to promote a more uniform distribution that improves performance over the population's eye tissue variance. Less than +/-30% irradiance variance over any +/- 10 nm band is recommended between 725 nm and 875 nm., with a levels tailing off <725 nm and >875 nm.*	**Partial Accept** There are no publicly available studies to specifically back these recommendations. The response to similar comments follows: The topic of wavelength characterization was considered at the workshop. The following points were discussed: • The primary motivation for the characterization is in the name of interoperability, following the NIST Special Publication (500-280) on *MobileID Device Best Practice Recommendations* and the draft version of the ISO/IEC 29794-6 which makes a specific recommendation

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				yet for all color camera pixels masked by a Bayer filter only the red pixels respond to the >90% content in NIR (700~835 nm), as the blue and green Bayer filters substantially block 700~830 nm. Therefore, as an example, despite that the <u>red</u> color pixels barely respond to <u>blue</u> illumination, the draft Special Publication virtually precludes supplying more than 10% blue light that has already been shown to increase overall performance when used at higher levels. This overly constraining percentage metric is counterproductive to promoting potentially the richest illumination mixture for iris imaging. Admittedly, the richest mixture of illumination for iris imaging has probably not been fully identified and documented, which the referenced paper states, but now is not the time to unwittingly preclude future illumination improvements by de facto standardization. Yet at the same time promoting the avoidance narrowband illumination is wise. • The recommended for changes to lines items 327, 328 and 329 follows: *The NIR broadband content between 700 and 900 nm is recommended to promote a more uniform distribution that improves performance over the population's eye tissue variance. Less than +/-30% irradiance variance over any +/- 10 nm band is recommended between 725 nm and 875 nm., with a levels tailing off <725 nm and >875 nm.* By limiting the lowest wavelength to 725 nm it promotes a falloff of irradiance as the NIR visibility becomes much		regarding operational wavelength of iris devices. However, it was acknowledged that there are not studies which back these specific guidelines. The current draft version does not use the characterization in formulating the qualification criteria. • It was noted that there are studies with some evidence showing that matching performance decreases with wavelength changes on the order of 100nm (e.g. Ngo et al 2009), with one algorithm but this is insufficient to make a specific quantitative recommendation. • Currently there are no comprehensive studies characterizing how matching performance depends on wavelength with the latest commercial matching algorithms, at least with a fine wavelength sampling within the range used by commercial iris cameras (700nm-900nm). Before such a quantitative recommendation is used in a qualification criteria, such a statement should be backed by such a study. A dataset has been identified which a preliminary study may be conducted. • It would be expensive and impractical to perform a data collection exploring all conceivable combinations

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				more apparent at the lower wavelengths, yet the lower wavelengths are also not necessarily marginally productive within the context of broadband content. Likewise, irradiance at wavelengths >875 nm is less meaningful to include within a specification because the variance levels >875 nm has far less impact to the system, especially as the sensor sensitivity falls off and negates its marginal contribution. Lastly concluding comments to no 2, by eliminating the content percentage metric in the draft, other potentially beneficial wavelengths like blue (400~500 m) will no longer be precluded.		of illumination profiles, between 700-900, considering multiple components with varying bandwidths. **Action:** Keep characterization as a measured quantity, but do not include as qualification criteria. Any wavelength characterization will not be made public. A study of wavelength interoperability that includes a number of commercial algorithms may be conducted.
ORG_G/ 3				Draft NIST Special Publication Line item heading 397, Section 3.3 Review of Specific Image Diagnostic Image Test Patterns, including Figure 4 provides an array of test patterns for characterizing the photonic and optical system performance for a set of two dimensional (2D) targets over an appropriate spatial frequency range. However, the human iris is a three dimensional (3D) target and there appears to be no attempt to include or account for the system performance including the third dimension, or iris pattern depth. Iris pattern depth in not inconsequential and should most definitely be included and accounted for in the system performance. To characterize the iris by 2D modeling at high spatial frequencies and yet not include and account for 3D iris texture is a glaring omission, especially for systems designed to leverage 3D iris texture. In the paper, 'Why Illuminant direction is fundamental to Texture analysis", M. J. Chantler exposes the effects of both illuminant direction and 3D texture. The proposed test target set of Figure 4 should add a 3D		**Reject** *(for this version)* Including 3D aspects of the iris into this test would add significant complexity to the manufacturing process. The editor acknowledges the possibility that feature contrast may be enhanced by the shadowing effects from off-axis illumination, however it has not been demonstrated that this is an important aspect of widely available iris capture devices. If a device manufacturer has concerns regarding the 2D nature of the IDQT as a source of bias, it should be identified and expressed to the testing operators so the possible bias is noted.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				texture/pattern that models some of the 3D texture characteristics of the iris. By including such a 3D target it would reveal in the test results of any combination effects of directional illuminators that improves iris performance by leveraging 3D texture. Visual examples of the 3D nature of the iris texture can be readily viewed at: http://www.gonioscopy.org/ http://www.surenmanvelyan.com/.		
ORG_G/ 4				Approximately one-third of the population wears eyeglasses. It would be highly relevant to include and account for system performance for eyeglass wearers. Or at a minimum assess and indicate whether a system is compatible with eyeglasses or not. Assessing actual system performance levels with eyeglasses is preferred.	Include an account for system performance with eyeglass wearers.	**Reject** It is acknowledged that eyeglasses can influence image quality, and that different devices may do a better job than others in mitigating the detrimental impact from eyeglasses. However, there are too many variables associated with eyeglasses and how they may fit on a face to make a simple, unbiased, effective performance test for eyeglass users. Considering the high rate of eyeglass use in the population, device performance related to eyeglasses will be handled in evaluation stages involving human subjects.
ORG_H/ 1				**TERMINOLOGY CONSISTANCY** The table below illustrates the current naming conventions and organization used in the First Public draft Test Plan. It should be possible to use one common name for the Processing, Metrics and the Measurements, and there should be a consistent number of them, and the order of presentation of these key elements. In most cases there	**ACTION**: **Consensus should be reached on the number of distinct measures, their identifying name and order of appearance in the plan**.	**Accept** The final version of the document will include a table that clarifies the measurements to be made, and which ones will be included to formulate the qualification criteria.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				is an obvious commonality or inferred similarity, which was used to assemble this table.		
ORG_H/ 2				**Star Target Selection** **Excerpt from 3.3.3 Gradient Contrast Star Patterns (60 and 120 segments)** 442 These targets are designed for direct measurement of the CTF using contrast targets with albedo 443 characteristics representative of those measured for the human iris. The 60 segment gradient 444 contrast star pattern covers frequencies ranging from **0.8 lp/mm to 3.5 lp/mm**. The 120 segment 445 pattern covers a higher range between 6.5 lp/mm and 1.6 lp/mm. There is a slight overlap in 446 coverage between the two targets to confirm results with separate physical targets for the level II 447 and level III criterion. In the workshop presentation (slide 25), the error in line 445 was corrected – "between 1.6 lp/mm and 6.5 lp/mm" and this correction needs to be carried over into the plan text. My comment/question is the ranges of lp/mm for the chosen number of segments. Since the evaluation is being performed at 1, 2 and 3 lp/mm, it would seem that only the 60 segment target is needed. I understand the desire to test devices which may have higher frequency capabilities (which justifies a target with more segments and a higher max lp/mm). BUT it seems that a different pair of numbers of segments would have been more beneficial. What would the ranges be if, for example, the numbers of segments were 40 and 90		**Reject** **Clarification:** The radial targets are chosen to provide high signal-to-noise measurements of MTF at spatial frequencies at 1, 2, and 3 lp/mm; and, to modulate the radial location of where specific frequencies are made. To note, the outer radial regions of the target contain the most area to make MTF measurements with star targets that result in measurements with more confidence.

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				(instead of 60 – 120)? Would both targets then allow for evaluation over the range of 1.0 to 3.0 lp/mm? (I believe it is imperative that the number of segments must be a factor of 360 to be feasible.) Were 60 and 120 chosen for any other specific reasons other than lp/mm range? It would seem that more overlap between targets would be more beneficial than 6.5 lp/mm which is way over the top evaluation level of 3 lp/mm.		
ORG_H/ 3				**Head Model Management** There was some discussion at the workshop on just how the head model design and manufacturing details would be managed. On one hand the information must to open to those needing the models to conduct independent testing (accredited labs). On the other hand, it may be necessary to NOT allow open access to camera developers who could tune to the test. At this point, I would recommend a written description of the Policy and technical descriptive documentation planned to manage the head model information. The solution should not be any form of sole-source supplier of models, nor should it be single-lab monopoly on testing.		**Accept** It is agreed that the best test would be to develop a policy to both facilitate the 3rd party testing integrity, while facilitating the development in industry to match the requirements of the US Government. The IDQT can be viewed as part of a requirements list, which requires knowledge of adherence to the test. There is an argument to make available some of the testing hardware to the vendors. These details are in development and may be included in the final document.
ORG_H/ 4				**Illumination Testing Procedures** In the text around lines 646-660, I recommend additional material to describe the layout of the illumination for the various levels. I think that the orientation of the "capture axis" relative to the illumination source may make a big difference. In real life, in direct sunlight, it may matter if the sun is overhead, 45 degrees from the side, or 45 degrees		**Accept** More details regarding how the ambient light test will be executed will be included in the final IDQT document. In addition, a characterization of the lighting environments for the basis of the testing will be included.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1 **Organization /Comment ID**	2 **Section/ line #**	(3) **Paragraph/ Figure/Table/ Note** (e.g. Table 1)	4 **Type of com-ment**[2]	5 **Comment (justification for change) by the organization**	(6) **Proposed change by the organization**	(7) **Proposed Editors Disposition**
				behind the camera. The recommendation is to at least document exactly what the configuration is. Beyond that, if practical, describe what real-life condition (s) this is intended to represent. I am not really recommending a whose suite of different illumination configurations.		
ORG_H/ 5				**Number of Images Captured in a Test** Excerpt from draft: 672 4.4 Application of Image Processing Algorithms 673 The result of a successful image capture process for a device in a single image per capture attempt 674 mode is a total of 530 images. I cannot reproduce this value. Recommend a small table to enumerate these. Table should contain which targets are imaged, how many rotations for each target, how many repetitions, which/how many illuminations, which coatings (and any other necessary variables needed to make this precise). Also included in this or another table could be images collected for: • Exposure time • Safety (Phototransistor and fiber spectrum) I'm sure something is wrong with my estimates below - but what? 6 targets x 4 rotations x 5 attempts x 2 eyes x 4 lighting levels = 960 (and this does not include coatings)		**Accept** For clarity, a table detailing the number of images used in the test will be included in the final IDQT document.
ORG_H/ 6				**Target Nomenclature** In Section 3.3 Figure 4 and section 3.3.n, there are different names used for some of the targets, and they are presented in different orders. Review this section and align to one title for each target. Then assure that all other references throughout the text are consistent with this terminology. EG: “Distortion Grid” or “Distortion Square Grid”?		**Accept** The target names will be made consistent throughout the document.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
ORG_H/ 7				**Capture Volume Evaluation** Excerpt from draft: 778 **Capture Volume Estimation** 779 The iris-like feature target is used for this exercise. Capture attempts are made throughout the 780 capture volume as claimed by the device manufacture. If important for a given application, the full 781 suite of image collection and analysis can be carried out. For the nominal test, captures attempts at 782 the boundaries of the capture volume are carried out with the iris texture target, with real time 783 feedback matching to a reference template. Discrepancies are from manufactures claims are noted in 784 the IDQT report. With regard to the testing procedure for this measure, there should be more detailed descriptions and possibly subsections that are related to the Mode. Capture volume testing for binocular types or those with mechanical aids will be very different from walk-through or stop-and-go type systems. In particular, for cameras with medium to large design capture volumes, it may not be sufficient to "capture attempts at the boundaries" but rather may warrant a series of steps in the vicinity of the boundary to allow determination of where the boundary really is. I think a binary (pass/fail, yes/no) at the boundary/corners would be insufficient.		**Accept** This is a good point, and in development tests this process of taking iterative steps around the manufacturer's recommended capture volume was practiced. The capture volume measurement procedure will be expanded upon in the final version.
ORG_I 1	General		ed	There are numerous examples of clumsy wording, missing or redundant words or phrases, etc. that have not been detailed here.	Give the document a careful proofreading to make sure the prose makes sense and is clear.	**Accept**
ORG_I 2	General		ed	Many of the sections have first subsections or	Give the first paragraph or subsection	**Accept**

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				paragraphs that are "floating" such that they cannot be uniquely referenced using section numbers. This applies to sections 2, 3, 3.8, 4.3, 4.4, and 5	a heading such as "Overview", "Introduction", etc. and a section number, i.e. 2.1, 3.1, 5.1. Renumber the subsequent paragraphs.	
ORG_I 3	Various		ed	An additional level of subsection numbers should be added to sections 2.6, 4.3.1 and 4.3.2 to facilitate references to subparagraphs	Add subsection numbering to 2.6, 4.3.1, and 4.3.2	**Accept**
ORG_I 4	2.3		ed	In last sentence "test" is redundant	Delete "test"	**Accept** Corrected
ORG_I 5	2.6, 4.4.2		te	Fabrication of targets and measurement for both MTF and CTF seems redundant since they are mathematically related.	Delete either the MTF or CTF measurements, or provide justification for including both.	**Accept** CTF will be used in the measurement process due to the simplification in creating the targets. However, the MTF is more easily conveyed from a designer's perspective. The corrections are small. In the final, an appendix will be included, giving the relationship between the CTF and the MTF
ORG_I 6	2.6		ed	In subsection on Pixel Scale, third sentence, description of iris dimensions should explicitly state what the values refer to.	change to read "Iris diameter typically ranges between 10.2 and 13.0 mm with an average of about 11.8 mm."	**Accept** Corrected
ORG_I 7	2.6		te	In subsection on Greyscale Gain Linearity, will the target with known NIR albedo regions provide uniform albedo over the 700-900 nm range? If not, will the illumination spectra affect these linearity measurements?	Clarify target properties.	**Accept** The characterization of the relevant optical properties of the ink used in the targets over the 700-900nm range will be included in the final draft.
ORG_I 8	2.6		te	In subsection on Greyscale Gain Linearity, no mention is made of the effects of nonuniform illumination, ambient reflections, or light scatter from the nose on the albedo measurements. It seems likely that the illumination distribution could be sufficiently uniform to assure accurate segmentation and texture encoding,	We would suggest a different target design that would have the four albedo values arranged in a cyclical pattern of local patches, like a checkerboard – that way measurements of a particular albedo	**Reject** The possibility of a field dependent gain pattern is handled in the IDQT test protocol by taking 4 different orientations of the quadrant pattern. This will allow each calibration region

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of comment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				operations that are designed to accommodate local intensity variation, but would have enough nonuniformity to negatively impact grayscale linearity measurements.	value would be distributed across the entire target. The gray value for a particular albedo value would be determined by averaging the gray values of all patches with that albedo, thereby eliminating the influence of nonuniform illumination, reflection, etc. The variance in observed local gray values for each albedo value would give a good indication of the uniformity of the illumination.	to be sampled in multiple places around the iris area, albeit on separate exposures. The quadrant pattern can also make an independent measure of the MTF using the ISO slanted edge method.
ORG_I 9	2.6		te	Measurement of exposure time does not appear to be relevant in systems that are not designed to accommodate subject motion.	Consider eliminating exposure time measurements for systems designed to capture stationary subjects.	**Partial Accept** The measurement for exposure time was included in the IDQT to put devices in the context of the Mobile ID Device Guideline which has recommended levels of exposure time for freezing subject motion. Because there are other ways of mitigating subject motion besides shortening exposure time this information is not used for any qualification criteria.
ORG_I 10	2.6		ed	In subsection on Capture Volume the description is somewhat unclear.	Change first sentence to read "The capture volume is the physical space within which the eye must be located for an iris capture device to produce an image that satisfies a qualification criteria." In the second sentence change to read "…without a subject eye present…"	**Accepted**
ORG_I 11	3.3		te	Some cameras may need to detect the outer iris boundary (the limbus) as part of their capture process.	Add an annular scleral region to the outside of each target. It does not	**Response (Accept):** The IDQT target pattern is mounted on a 3D printed

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				The targets as described do not appear to provide any scleral region outside the iris.	need to be very wide – perhaps 25% of the iris diameter (i.e. 3 mm).	"eyeball" which has optical properties in line with those of the human sclera at observational wavelengths around 800nm.
ORG_I 12	3.3.1, 4.4.2	Figure 4	te	The edges in the Quadrant Pattern appear to make an angle of about 28 degrees to the vertical or horizontal axis. Most slant-edge MTF software, such as that available for ImageJ, appears to be optimized for edges that make an angle of about 5.8 degrees with the vertical or horizontal axis, as do those in the ISO 12233 target.	Modify the Quadrant Pattern accordingly or verify that the software used will work with the angles used.	**Accept** Text will be added to explicitly mention the 5.8 degree angle horizontal and vertical.
ORG_I 13	3.3.1, 4.4.2	Figure 4	te	No mention is made of the desired intensity values on each side of the slant edge. In our experience it is important that the "black" side of the edge have an intensity greater than zero and that the "bright" side have an intensity less than the saturation level of the camera to assure accurate MTF measurements.	Add this specification to the MTF measurement set-up unless it can be verified that the software used does not require this constraint.	**Partial Accept** **Clarification**: Reflectivity of the different patches was chosen to represent the range found in the human iris. Zeros or saturated pixels. in the regions of delivered images, would be noted in the linearity test; and, this would likely result in a test failure on the MTF and/or texture target tests.
ORG_I 14	3.7	Table 1	te	This table assumes that indoor operation with sunlight through glass is equivalent to outdoor operation in the shade. Has this been verified through actual measurements? Outdoor operation in shade is important because it affords subjects the opportunity to open their eyes wide without excessive discomfort. It may be appropriate to define outdoor shaded operation as a separate ambient light scenario.	Investigate or clarify and consider adding an additional ambient light scenario for outdoor in shade.	**Accept** The editor will investigate the differences between outdoor versus existing indoor sunlight through glass.
ORG_I 15	3.8.1		ed	"manufacture" is a verb – what is meant here is the noun, which is "manufacturer".	Change "manufacture" to "manufacturer".	**Accept**
ORG_I 16	4.3		te	It appears that not much thought has been given to the software used to capture test images, and for most if	Include at least a brief functional specification for capture software to	**Accept** Different devices vary in how they

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment**[2]	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				not all cameras, custom software will probably be needed to generate a particular image size, provide for manual (forced) capture, control illumination, encode capture information in the output filename, etc.	be provided by the camera vendor.	deliver images intended for biometric matching and in the availability of interface SDKs and/or demonstration applications. There is intention in not providing any specifics on capture software for a device interface to enable consideration of devices that may not necessarily have a polished interface. Specifics needed to conduct the test would be handled via communication with the device manufacturer. The bare requirement is that the device will make available collected images intended for use in iris biometrics. For example, the test does not presently have requirements on image dimension or bit depth. The point is well taken that there is a need for the definition of what is meant by a "device" in addition to an explicit definition of the requirements, from the device manufacture in order to execute the test.
ORG_I 17	4.3.2		te	In "Collection Procedure for Ambient Light Qualification" no mention is made of cameras that have particular operational configurations for outdoor use. For example, some cameras have hoods or baffles designed to block ambient light when used in high ambient light environments. Cameras that have such devices should be tested with them in place.	Acknowledge the use of devices designed to block ambient light and include instructions to use such devices when doing collection for ambient light qualification, in accordance with vendor recommendations.	**Accept** Although not explicitly stated, it was intended that devices with baffles would be tested with baffles in place, or per manufacturer instructions. Wording will be added to the document to ensure this clear.
ORG_I 18	4.4.3		te	This section is somewhat unclear. Is the "mask defined by where the recorded signal strength is below a	Consider eliminating the use of a signal quality mask. The target images	**Reject** **Clarification**: A signal quality mask (also

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				threshold" the same as the "signal quality mask" of 5.1? Where does the "pristine reference template" come from? Does it have any masked elements? How is the threshold for pixel masking determined? In the next-to-last sentence of the first paragraph, "…and the percentage of the 20 values of the total area which passes the mask filter" is unclear.	have no eyelids or eyelashes so no masking based on the input image is necessary. If we assume that the "pristine reference template" is generated from an image with the best achievable iris SNR, then the best indicator of a particular camera's SNR is the HD that it achieves when matched against this "pristine reference template". Otherwise you must define how the mask threshold is determined. In at least some template generators the threshold is dynamically set for each image so as to produce a certain percentage of masked elements in the template. But this would guarantee that every template has a fixed percentage of its elements masked, which would certainly defeat the purpose of the measurements described in 5.1. As for the next-to-last sentence of the first paragraph, it could be changed to read "…and, for each of the 20 images, the percentage of the total area that passes the mask filter" if this accurately states what is intended.	known as a "fragile bit" mask) is included in the score generation to test the aspects of matching algorithms with known benefits and incorporated into commercial products. The specifics of the mask definition are sequestered at this point. What can be said is that the mask accommodates the SNR scale per device, which can vary arbitrarily by things like gain settings by a normalization. This normalization is based on baseline SNRs measurements from the uniform patches of the quadrant target. The method has been validated over a variety of image outputs and is allowed to go "full pass" for instances such as when the SNR is very low compared to the dynamic range measure over the iris albedo range.
ORG_I 19	5.1			In each description of the three levels, the third sentence reads "Each Hamming distance is only valid if more than 90% of the iris area passes the signal quality mask relative to the reference template mask." This is not clear.	Although it is not clear what is intended, one interpretation would be better expressed as "Each Hamming distance is valid only if at least 90% of the bits in the logical AND	**Partial Accept** **Clarification**: The definition of the pristine template (i.e. generated the original digital form of the signal) has a nominal definition of the pristine

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				See also ORG_I 18.	of the probe mask and the reference mask indicate valid results in the logical XOR of the corresponding template bits." If the reference template were truly "pristine" it would have NO masked elements (the mask would contain no zeros), and the logical AND of the probe and reference masks would just reflect the "usable iris area" of the probe template.	template which results in under 2% of the template area under a mask. The suggested clarification, although worded differently, is equivalent to the score formation method used. The suggested wording will be adopted in the text.
ORG_I 20	Appendix A			It may not be clear to the reader why the specific qualification conclusions are justified. It would be helpful if more details were provided on the basis for each qualification decision.	Provide details on justification for the qualification levels, i.e. whether they are based on iris texture scores or ambient lighting noise scores and why.	**Partial Accept** Some justification to the criteria was presented at the BCC, but not included in the first draft of the IDQT document. There will be further justification added to the next draft of the documents.
ORG_J 1			g	For iris images, matching performance is of paramount importance, and image characteristics which do not significantly affect biometric performance should not be strongly weighted in the final IDQT result. This is obviously appreciated by the authors of the document, but could do with a little stronger emphasis. An obvious counter-argument to this is that there is no predicting what features will be important in future algorithms. However the basic iris coding techniques have been stable for a relatively long time, and it is probably not likely that significant changes will appear in the short term.		**Accept** Text will be added to emphasize the point that the qualification criteria are only included on items which have proven correlation to aspects of biometric performance.
ORG_J 2				On the exposure time measurement, some careful thought needs to be applied to the arrangement of light sources, so that cameras with rolling shutters can be		**Accept** This is a good point. It was assumed that these effects would be averaged

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				properly characterized,		out over the multiple exposures taken but this assumption was not validated. The possibility of faulty measurements for systems with rolling shutters will be investigated.
ORG_J 3				On duel wavelength illumination. We have experimented with multiple wavelengths, and have noticed no significant difference in biometric performance between mixed 780nm+850nm wavelength illumination and single 850nm wavelength illumination. It is possible that the 780nm illumination may improve iris sclera contrast, and therefore improve segmentation, but we have not been able to to show this in our experiments. Due to the several patents in this area NIST should weigh carefully the possible biometric benefits, against the potential for unnecessarily narrowing the supplier base.		**Partial Accept** This study would need to be repeated by a 3rd party entity or NIST. Any study offered as a contribution is welcome as consideration to initiate a validation effort.
ORG_J 4				ORG_J devices use reflections from the cornea as an intrinsic part of the capture process. This can be considered a first order liveness test, but it cannot be turned off, since the capture process relies on these reflections being present. The experimental design described in this document appears to account for this possibility, so we do not anticipate any problems. For this reason at least, it is important that the test targets mimic real eyes as much as possible.		**Noted** We have taken care in replication of the reflective properties of the iris; however, if there are any suspected capture failures, the IDQT will communicate these to vendors to identify if the source of the failure is that the test target does not encompass the signal requirements of the device.
ORG_K 1	all		Ge	The document does not state who the intended audience is (or is not). SNR, albedo, etc. imply device manufacturer engineers and scientists, but references to "down selection" imply USG systems engineers and acquisitions staff. Following the July 9th workshop, ORG_K has a clearer understanding of the document's	Consider expanding the Introduction to include Intended Audience(s) and stating clearly and concisely for whom the document was (and was not) written.	**Accept** The draft is in the process of being edited to make it more accessible to a wider audience.

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
				intended audience, but the uninformed may be protected from themselves if the document includes a clear, concise statement of for whom it is (and is not) intended.		
ORG_K 2	1		Te	Section 1, Introduction does not clearly explain that when IQDT is complete there basically will be a products list based on Sections 5.1-5.4. ¶3 mentions application requirements, but does not explain (or point to references that explain) where or how the reader is supposed to gather device requirements to help cull the field between market survey and device qualification.	Edit for brevity and clarity. See #1.	**Partial Accept** The resulting products from the IDQT are not solidified at this point; however, the intended use could (and will be) expanded upon in the introduction.
ORG_K 3	1	Footnote 1	te	Differentiating between 19794-6:2011 and 19794-6:2005, which permits polar format, is too important to relegate to a footnote. JMS did not see this footnote on the first reading and only found it when searching the text for 19794-6 to write a comment about undated references to 19794-6, JPEG, etc.	Consider inserting “Normative References” – or – Promoting the text from the footnote into the document body proper.	**Accept** Text from the footnote will be added to the document body.
ORG_K 4	2		Ge	The document is full of good, useful, meaningful information that is not altogether well organized, e.g., much of Section 2, Test Overview and Scope belongs in Section 1, Introduction (or should be split into separate sections). Much of it seems repetitive, what the document *is not* as opposed to what it is, and may better suited to an annex than the document body.	“Test Overview” through and including 2.5, Levels of qualification: Move to the Introduction – or – Split from Scope Renumber 2.6, In-Scope Measurements as appropriate Retitle 2.6, In-Scope Measurements to Scope and add a sentence/paragraph, if necessary Insert 2.6.1, In-Scope above “The following subsections…” Renumber 2.7, Out-of-scope to 2.6.2, Out-of-scope	**Partial Accept** Extensive reorganization of the document is being considered for the next draft. The suggested reorganization will be considered in the context of the content additions stemming from addressing comments.
ORG_K 5	2		ed	Section 2 is inconsistent with Sections 3 and 4 in that it does not contain 3rd level headings, e.g., Spatial Frequency Response, not 2.6.1, Spatial Frequency Response.	Add 3rd level headings in Sections 2.1, 2.5, and 2.6	**Accept** Inconsistencies, such as those noted, will be considered (and/or corrected) in

Title: Comments on IDQT draft (pre workshop)

Date: May 14, 2013	Document: IDQT Draft (pre-workshop)

1	2	(3)	4	5	(6)	(7)
Organization /Comment ID	**Section/ line #**	**Paragraph/ Figure/Table/ Note** (e.g. Table 1)	**Type of com-ment[2]**	**Comment (justification for change) by the organization**	**Proposed change by the organization**	**Proposed Editors Disposition**
						future revisions.
ORG_K 6	2.3		Te	PNG, JPEG 2000, and TIFF are referenced only once and are not expanded from their acronyms nor referenced properly to point to a specific or undated version of their respective ISO standards. Reference image formats, e.g., ISO/IEC 19794-6, PNG and JPEG2000, without the usual boilerplate text that all undated references explicitly mean the most recent version.	In conjunction with #3: Consider inserting “Normative References” – or – Inserting complete references for PNG, JPEG 2000, and TIFF	**Accept** Necessary references will be considered (and/or corrected) in future revisions.
ORG_K 7	2.3		Te	References undated ISO/IEC 19794-6 Is polar format from 2005 allowed or excluded?	Consider adding text that clearly states ISO/IEC 19794-6:2005 polar format is not permitted.	**Accept** Formats are not permitted, which will be explicitly stated.
ORG_K 8	2.6 4.4.x		ge	Sections 2.6 and 4.4.x contain redundant information re: many of the test measurements.	Consider adding a section (or appendix) for definitions and moving much/all of the content there and limiting discussion in Section 2 to what is in/out of scope and Section 4 to how something is measured, not what that something is.	**Accept** This is a good suggestion and will be included in the next draft.
ORG_K 9	4.4.x		Ed	Sections 4.4.x re-hash most/much of the content of 2.6 and lose ***how*** each element is measured in a description of ***what*** each element is.	Edit for brevity and describe only ***how*** a measurement is taken, not ***what*** the measurement is.	**Accept**